

DOCUMENT RESUME

ED 132 206

TH 005 943

AUTHOR Garvin, Alfred D.
TITLE Three Approximations of Standard Error of Measurement: An Empirical Approach.
PUB DATE [Sep 76]
NOTE 10p.; Paper presented at the Annual Meeting of the American Psychological Association (84th, Washington, D.C., September 1976)
EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS *Standard Error of Measurement; *Statistical Analysis; Test Reliability

ABSTRACT

Three successively simpler formulas for approximating the standard error of measurement were derived by applying successively more simplifying assumptions to the standard formula based on the standard deviation and the Kuder-Richardson formula 20 estimate of reliability. The accuracy of each of these three formulas, with respect to the standard formula, was assessed empirically in a sample of 60 sets of test results, representing a wide variety of test characteristics. It was found that all three formulas provided an essentially unbiased approximation of the value given by the standard formula: the mean error was +.04 of a score point for each formula. The standard deviations of these errors were .09, .10, and .25 of a score point for the three formulas respectively. It was concluded that any of these three formulas was accurate enough for most practical purposes. (Author/RC)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

Three Approximations of Standard Error of Measurement

An Empirical Approach

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

Alfred D. Garvin

University of Cincinnati

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ABSTRACT

Three successively simpler formulas for approximating the standard error of measurement (SEM) were derived by applying successively more simplifying assumptions to the standard formula based on the standard deviation and the Kuder-Richardson formula 20 estimate of reliability.

These formulas are as follows:

$$SEM_1 = \frac{\sqrt{N \sum T - \sum T^2}}{N}$$

where T is the number of testees who answered a given item correctly and N is the number of testees;

$$SEM_2 = .9 \sqrt{\frac{\bar{X}(k - \bar{X})}{k}}$$

where $\bar{X}$ is the mean of the scores earned and k is the number of items; and

$$SEM_3 = .14 \sqrt{k}$$

The accuracy of each of these three formulas, with respect to the standard formula, was assessed empirically in a sample of 60 sets of test results, representing a wide variety of test characteristics. It was found that all three formulas provide an essentially unbiased approximation of the value given by the standard formula: the mean error was

+ .04 of a score point for each formula. The standard deviations of these errors were .09, .10, and .25 of a score point for formulas 1, 2, and 3, respectively. It was concluded that any of these three formulas is accurate enough for most practical purposes.

Three Approximations of Standard Error of Measurement:

An Empirical Approach

Alfred D. Garvin

College of Education

University of Cincinnati

Cincinnati, Ohio 45221

The standard error of measurement (SEM) is potentially a very useful test statistic for classroom teachers-testers-decision makers. For example, it can provide them with a defensible basis for setting a score band around a "cutting score" within which they might impose probation on students or require their reexamination. The practical problem is that, even with the advent of inexpensive electronic calculators, the computations involved in the conventional formula for SEM are so tedious that most such teachers are discouraged from undertaking the task. Instead, they either set an arbitrary score band within which they exercise some administrative option or, more commonly, they set a single cutting score with no provision for a second chance or a second look.

It should be remembered that the familiar definitional formula for SEM is deceptively simple:

$$SEM = S \sqrt{1 - r_{xx}} \quad (1)$$

where S is the estimate of the standard deviation of the scores of a population of testers and r_{xx} is some estimate of the reliability of the test used. However, when we substitute the computational formulas for S and,

say, the K-R (20) version of r_{xx} in formula (1) the following formidable computational task results:

$$SEM = \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n(n-1)}} \sqrt{1 - \frac{k}{k-1} \left[1 - \frac{n \sum T - \sum T^2}{n \sum X^2 - (\sum X)^2} \right]} \quad (2)$$

where: n is the number of testees,

k is the number of items (each scored 1 or 0)

X is the score earned by a given testee, and

T is the number of testees who answered a given item correctly.

For the purpose of this paper, we shall term the method of computing SEM shown in formula (2) the exact method. It is beyond the scope of this paper to argue whether or not it is the correct method from a conceptual point of view.

To summarize the problem, despite the potential value of an estimate of SEM to classroom teachers-testers-decision makers, too few of them are willing to undertake the complex computations involved in the so-called exact method. What is needed, then, is some usefully accurate, much simpler method of approximating SEM.

Three Approximations of SEM

If the terms $n - 1$ and $k - 1$ in formula (2) were replaced by n and k , respectively, formula (2) could be reduced to:

$$SEM = \frac{\sqrt{n \sum T - \sum T^2}}{n} \quad (3)$$

These two algorithmic liberties are reasonable on two accounts. First, and more obviously, if n and k are relatively large, as they are likely to be in any "important" testing situation, the effects of these two substitutions would be relatively small. Second, and less obviously, in any testing situation, the effects of these two substitutions tend to offset each other.

Formula (3) requires substantially less computation than the so-called exact method while promising a very accurate approximation of SEM under most foreseeable test conditions.

It will be remembered that the exact method is based on the K-R (20) version of r_{xx} . If the K-R (21) version were used instead, together with the substitution of n and k for $n - 1$ and $k - 1$, such an approximation of the exact method could be reduced to:

$$SEM = \sqrt{\frac{\bar{X}(k - \bar{X})}{k}} \quad (4)$$

It might be noted that formula (4) is superficially similar to one given by Lord (1957) for the SEM of a given score. However, the concept of SEM dealt with in this paper is the conventional one of a global test statistic applicable to any score in a score distribution.

To return to the assumptions involved in formula (4), the K-R (21) formula for r_{xx} assumes that all items of the test involved have the same difficulty for the testees involved. In the notation of this paper, it assumes that all T s are equal. The more this assumption is violated in

test data the more the K-R (21) r_{xx} underestimates the K-R (20) r_{xx} and the more any approximation of SEM based on the K-R (21) r_{xx} overestimates the exact SEM. To compensate for this effect of the inevitable violations of the assumption of equal item difficulties, a constant empirically derived from pilot studies was introduced into formula (4), viz:

$$SEM = .9 \sqrt{\frac{\bar{X}(k - \bar{X})}{k}} \quad (4A)$$

Formula (4A) requires even less computation than formula (3) while promising a usefully accurate approximation of SEM under most foreseeable test conditions.

Referring to formula (4A), if the further assumption is made that $\bar{X}$ is about 73 percent of k , formula (4A) could be reduced to:

$$SEM = .4 \sqrt{k} \quad (5)$$

Formula (5) is surely the ultimate in computational simplicity and yet the simplifying assumptions on which it is based are all reasonable in the light of ~~empirical~~ experience.

To summarize, three successively simpler formulas for approximating SEM were derived by adopting more and more empirically suggested assumptions and reducing the formula for the so-called exact method accordingly. It remained to be seen how accurate each of these three approximations of SEM was in an empirical evaluation.

Procedure

Sixty sets of results from teacher-made, objective achievement tests were randomly selected from among a much larger number submitted to this investigator as projects by students in his Tests and Measurement course. Each test had been written by a different teacher or team of teachers and each test had been taken by a different group of testees. In all, 1792 testees and 1998 test items were involved in these 60 testing situations. The most common item format was four-choice multiple choice, with smaller blocks of true-false, completion, and matching items sometimes added in that order of frequency. In content area and level, these tests ranged from kindergarten reading readiness to graduate psychiatric nursing. The distributions of relevant test and group characteristics over the 60 tests are summarized in Table 1.

TABLE 1

Distributions of Test and Group Characteristics

	k	K-R (20) r_{xx}	SEM	n
Maximum	100	.93	3.78	180
Median	28	.67	2.12	26
Minimum	13	.22	1.42	10

It seems fair to say that this sample of 60 sets of test results was representative of the full range of those that might be encountered in classroom testing generally.

For each of these 60 sets of test results an exact SEM was computed according to formula (2) and approximations of SEM were computed according to formulas (3), (4A), and (5). Then, for each of these 60 sets of test results the error of each of the three approximations of SEM from the exact SEM was computed. Finally, these errors were summarized over the 60 tests for each of the three approximations.

Results

Table 2 presents the mean, standard deviation, and maximum value of the errors of approximation over the 60 sets of test results for each of the three approximations of SEM.

TABLE 2

Mean, Standard Deviation, and Maximum Value
of Errors in Three Approximations of SEM

	Formulas		
	(3)	(4A)	(5)
Mean	+ .04	+ .04	+ .04
S.D.	.09	.10	.25
Max.	+ .37	+ .34	- .66

No one approximate method was consistently more accurate than the other two over all tests. Each method seemed to be most vulnerable to a different set of test-group characteristics.

Conclusions

It seems reasonable to conclude that each of these three methods of approximating SEM provides a usefully accurate, essentially unbiased approximation of the exact SEM over the full range of typical classroom testing situations. As might be expected, the simplest formula, (5), yields the least consistent approximations but, even here, over two-thirds of the values obtained fell within one-quarter of a score point of the exact value. When it is remembered that the SEM obtained is rounded to the nearest whole number for most practical applications it can be concluded that any of these three methods makes the potential advantages of the SEM readily available to any classroom teacher-tester-decision maker who is concerned with defensible evaluation practices.

A Final Comment

It would be ungenerous, indeed, to conclude this paper without a final comment on Lord's (1957) classic article, "Do Tests of the Same Length Have the Same Standard Error of Measurement?". While this paper dealt with a quite different concept of SEM and reasoned from a quite different set of assumptions, the data adduced here through formula (5) do, indeed, confirm Lord's (1959) equally classic answer to his own question: In most classroom testing situations, "Tests of the same length do have the same standard error of measurement."

References

- Lord, F. M. Do tests of the same length have the same standard error of measurement? Educational and Psychological Measurement, 1957, 17, 510-521.
- Lord, F. M. Tests of the same length do have the same standard error of measurement. Educational and Psychological Measurement, 1959, 19, 233-239.